

Pending applications - Search notes are with the Examiner (Intergovernmental)

W P E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Fri Jun 16 15:51:26 2000; MasPar time 3.15 Seconds
Tabular output not generated. 36.626 Million cell updates/sec

Title: >US-08-905-046-2
Description: (1-8) from US08905046.pep
Perfect Score: 48
Sequence: 1 SAVALTYS 8

Scoring table:
PAM 150
Gap 15

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1
Statistics: Mean 14.380; Variance 37.613; scale 0.382

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description	Pred. No.
1	48	100.0	36	4 PCT-US96-0 Sequence 1, Applicatio	1.83e+01
2	48	100.0	37	4 PCT-US96-0 Sequence 15, Applicati	1.83e+01
3	48	100.0	37	2 US-08-460- Sequence 4, Applicatio	1.83e+01
4	48	100.0	37	2 US-08-460- Sequence 3, Applicatio	1.83e+01
5	44	91.7	36	4 PCT-US96-0 Sequence 2, Applicatio	5.85e+01
6	44	91.7	36	2 US-08-460- Sequence 2, Applicatio	5.85e+01
7	43	89.6	37	4 PCT-US96-0 Sequence 4, Applicatio	7.78e+01
8	43	89.6	37	2 US-08-460- Sequence 5, Applicatio	7.78e+01
9	43	89.6	148	4 PCT-US96-0 Sequence 10, Applicati	7.78e+01
10	43	89.6	171	2 US-08-483- Sequence 10, Applicati	7.78e+01
11	42	87.5	37	2 US-08-460- Sequence 6, Applicatio	1.03e+02
12	42	87.5	37	4 PCT-US96-0 Sequence 5, Applicatio	1.03e+02
13	42	87.5	117	4 PCT-US96-0 Sequence 11, Applicati	1.03e+02
14	40	83.3	36	4 PCT-US96-0 Sequence 6, Applicatio	1.81e+02
15	40	83.3	36	2 US-08-460- Sequence 7, Applicatio	1.81e+02
16	40	83.3	40	4 PCT-US96-0 Sequence 12, Applicati	1.81e+02
17	40	83.3	415	1 US-08-110- Sequence 2, Applicatio	1.81e+02
18	40	83.3	415	1 US-08-110- Sequence 6, Applicatio	1.81e+02
19	40	83.3	3218	1 US-08-764- Sequence 27, Applicati	1.81e+02
20	39	81.3	84	2 US-08-353- Sequence 80, Applicati	2.39e+02
21	39	81.3	375	2 US-08-871- Sequence 2, Applicatio	2.39e+02
22	39	81.3	375	2 US-08-819- Sequence 4, Applicatio	2.39e+02
23	38	79.2	22	2 US-08-525- Sequence 23, Applicati	3.15e+02

24	38	79.2	37	4 PCT-US96-0 Sequence 3, Applicatio	3.15e+02
25	38	79.2	51	4 PCT-US96-0 Sequence 13, Applicati	3.15e+02
26	38	79.2	147	4 PCT-US96-0 Sequence 9, Applicatio	3.15e+02
27	38	79.2	170	2 US-08-483- Sequence 11, Applicati	3.15e+02
28	38	79.2	170	1 US-08-460- Sequence 3, Applicatio	3.15e+02
29	38	79.2	178	3 US-09-040- Sequence 2, Applicatio	3.15e+02
30	38	79.2	178	3 US-09-081- Sequence 2, Applicatio	3.15e+02
31	38	79.2	320	1 US-08-555- Sequence 17, Applicati	3.15e+02
32	38	79.2	437	2 US-08-473- Sequence 9, Applicatio	3.15e+02
33	38	79.2	437	2 US-08-477- Sequence 9, Applicatio	3.15e+02
34	38	79.2	437	1 US-08-277- Sequence 6, Applicatio	3.15e+02
35	38	79.2	560	3 US-08-808- Sequence 18, Applicati	3.15e+02
36	38	79.2	560	3 US-08-808- Sequence 18, Applicati	3.15e+02
37	38	79.2	560	3 US-09-050- Sequence 18, Applicati	3.15e+02
38	38	79.2	670	2 US-08-477- Sequence 11, Applicati	3.15e+02
39	38	79.2	670	2 US-08-473- Sequence 11, Applicati	3.15e+02
40	37	77.1	251	4 PCT-US95-1 Sequence 1, Applicatio	4.13e+02
41	37	77.1	330	2 US-08-871- Sequence 18, Applicati	4.13e+02
42	37	77.1	347	1 US-08-118- Sequence 47, Applicati	4.13e+02
43	37	77.1	386	1 US-08-319- Sequence 14, Applicati	4.13e+02
44	37	77.1	393	1 US-07-629- Sequence 3, Applicatio	4.13e+02
45	37	77.1	398	2 US-08-288- Sequence 1, Applicatio	4.13e+02

ALIGNMENTS

RESULT 1					
ID	PCT-US96-08730-1	STANDARD:	PRT:	36 AA.	
XX	xxxxxx				
AC					
DT					
XX					

Sequence 1, Application PC/TUS9608730
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick
APPLICANT: Anderson, Jeffrey
TITLE OF INVENTION: Methods of Raising Antibodies Against E.
TITLE OF INVENTION: Coll of the Family CSF-CFA./1
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glenna Hendricks
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: USA
ZIP: 22031

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08730
FILING DATE: 03-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: PCT/US96/08730
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 591-4470
TELEFAX: (703) 591-4428
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide

RESULT	3	STANDARD;	PRT;	37 AA.
ID	US-08-460-617-4			
XX	xxxxxx			
AC				
XX				
DT				
XX				
DE				
XX	Sequence 4, Application US/08450617			
DE				
XX	Sequence 4, Application US/08450617			
CC	Patent No. 5914114			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Cassels, Frederick J			
CC	TITLE OF INVENTION: Method Of Raising Antibodies Against E.			
CC	TITLE OF INVENTION: Coll			
CC	NUMBER OF SEQUENCES: 9			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: Hendricks and Associates			
CC	STREET: 9669 A Main Street			
CC	CITY: Fairfax			
CC	STATE: VA			
CC	COUNTRY: US			
CC	ZIP: 22031			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patent In Release #1.0, Version #1.25			
CC	TITLE OF INVENTION: Coll of the Family CSF-CFA./1			
CC	NUMBER OF SEQUENCES: 15			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: Glenna Hendricks			
CC	STREET: P.O. Box 2509			
CC	CITY: Fairfax			
CC	STATE: VA			
CC	COUNTRY: USA			
CC	ZIP: 22031			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patent In Release #1.0, Version #1.25			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: PCT/US96/08730			
CC	FILING DATE: 03-JUN-1996			
CC	CLASSIFICATION:			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: Hendricks, Glenna			
CC	REGISTRATION NUMBER: 32,535			
CC	REFERENCE/DOCKET NUMBER: PCT/US96/08730			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: (703) 591-4470			
CC	TELEFAX: (703) 591-4428			
CC	INFORMATION FOR SEQ ID NO: 15:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 37 amino acids			
CC	TYPE: amino acid			
CC	STRANDEDNESS: single			
CC	TOPOLOGY: unknown			
CC	MOLECULE TYPE: peptide			
CC	HYPOTHETICAL: NO			
CC	ANTI-SENSE: NO			
CC	FRAGMENT TYPE: internal			
CC	SEQUENCE 37 AA; 3732 MW; 7516 CN;			
DB	27 SAVALTYS 34			
QY	1 SAVALTYS 8			
RESULT	4			
ID	US-08-460-617-3			
XX	xxxxxx			
AC				
XX				
DT				
XX				
DE				
XX	Sequence 3, Application US/08450617			
CC	Query Match			
CC	Best Local Similarity 100.0%; Pred. No. 1.83e+01;			
CC	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
DB	28 SAVALTYS 35			
QY	1 SAVALTYS 8			
RESULT	2			
ID	PCT-US96-08730-15			
XX	xxxxxx			
AC				
XX				
DT				
XX				
DE				
XX	Sequence 15, Application PC/TUS9608730			
CC	Sequence 15, Application PC/TUS9608730			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Cassels, Frederick			
CC	APPLICANT: Anderson, Jeffrey			
CC	APPLICANT: Carter, John Mark			
CC	TITLE OF INVENTION: Methods of Raising Antibodies Against E.			
CC	TITLE OF INVENTION: Coll of the Family CSF-CFA./1			

Sequence 3, Application US/08460617
Patent No. 5914114
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick J
TITLE OF INVENTION: Method of Raising Antibodies Against E.
COLI
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hendricks and Associates
STREET: 9669 A Main Street
CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,617
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: Cas460
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-425-4250
TELEFAX: 703-425-2767
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: E.coli

SEQUENCE 37 AA; 3732 MW; 7516 CN;
Query Match 100.0%; Score 48; DB 2; Length 37;
Best Local Similarity 100.0%; Pred.No.1.83e+01;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

28 SAVALTYS 35
|||||
1 SAVALTYS 8

QY

RESULT 5
ID PCT-US96-08730-2 STANDARD; PRT; 36 AA.
AC xxxxxx
XX
DT
DE
XX
XX
CC

Sequence 2, Application PC/TUS9608730
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick
APPLICANT: Anderson, Jeffrey
APPLICANT: Carter, John Mark
TITLE OF INVENTION: Methods of Raising Antibodies Against E.
COLI
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glenna Hendricks
STREET: P.O. Box 2509

CC

CC	CITY: Fairfax
CC	STATE: VA
CC	COUNTRY: USA
CC	ZIP: 22031
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/US96/08730
CC	FILING DATE: 03-JUN-1996
CC	CLASSIFICATION:
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Hendricks, Glenna
CC	REGISTRATION NUMBER: 32,535
CC	REFERENCE/DOCKET NUMBER: PCT/US96/08730
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: (703) 591-4470
CC	TELEFAX: (703) 591-4428
CC	INFORMATION FOR SEQ ID NO: 2:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 36 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: unknown
CC	MOLECULE TYPE: peptide
CC	HYPOTHETICAL: NO
CC	ANTI-SENSE: NO
CC	FRAGMENT TYPE: internal
SQ	SEQUENCE 36 AA; 3629 MW; 7132 CN;
	Query Match 91.7%; Score 44; DB 4; Length 36;
	Best Local Similarity 75.0%; Pred. No. 5,85e+01;
	Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Dd	27 ASVALTYS 34 ::
QY	1 SAVALTYS 8
RESULT 6	
ID US-08-460-617-2	STANDARD; PRT: 36 AA.
XX xxxxxx	
AC	
XX	
XX	
DT	
XX	
XX	
DE	Sequence 2, Application US/08460617
XX	
CC	Sequence 2, Application US/08460617
CC	Patent No. 5914114
CC	GENERAL INFORMATION:
CC	APPLICANT: Cassels, Frederick J
CC	TITLE OF INVENTION: Method of Raising Antibodies Against E.
CC	TITLE OF INVENTION: Coil
CC	NUMBER OF SEQUENCES: 9
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: Hendricks and Associates
CC	STREET: 9669 A Main Street
CC	CITY: Fairfax
CC	STATE: VA
CC	COUNTRY: US
CC	ZIP: 22031
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/460.617
CC	FILING DATE: 02-JUN-1995
CC	CLASSIFICATION: 42A

Cc		CITY: Fairfax
Cc		STATE: VA
Cc		COUNTRY: USA
Cc		ZIP: 22031
Cc		COMPUTER READABLE FORM:
Cc		MEDIUM TYPE: Floppy disk
Cc		COMPUTER: IBM PC compatible
Cc		OPERATING SYSTEM: PC-DOS/MS-DOS
Cc		SOFTWARE: PatentIn Release #1.0, Version #1.25
Cc		CURRENT APPLICATION NUMBER: PCT/US96/08730
Cc		FILING DATE: 03-JUN-1996
Cc		CLASSIFICATION:
Cc		ATTORNEY/AGENT INFORMATION:
Cc		NAME: Hendricks, Glenna
Cc		REGISTRATION NUMBER: 32,535
Cc		REFERENCE/POCKET NUMBER: PCT/US96/08730
Cc		TELECOMMUNICATION INFORMATION:
Cc		TELEPHONE: (703) 591-4470
Cc		TELEFAX: (703) 591-4428
Cc		INFORMATION FOR SEQ ID NO: 2:
Cc		SEQUENCE CHARACTERISTICS:
Cc		LENGTH: 36 amino acids
Cc		TYPE: amino acid
Cc		STRANDEDNESS: single
Cc		TOPOLOGY: unknown
Cc		MOLECULE TYPE: peptide
Cc		HYPOTHETICAL: NO
Cc		ANTI-SENSE: NO
Cc		FRAGMENT TYPE: internal
SQ		SEQUENCE 36 AA; 3629 MW; 7132 CN;
		Query Match 91.7%; Score 44; DB 4; Length 36;
		Best Local Similarity 75.0%; Pred. No. 5,85e+01;
		Matches 6; Conservative 2; Mismatches 0; Indels 0;
Dd	27	ASVALTYS 34
	:	:
Qy	1	SAVALTYS 8
RESULT	6	
ID	US-08-460-617-2	STANDARD; PRY: 36 AA.
XX	xxxxxx	
AC		
DT		
DE	Sequence 2, Application US/08460617	
XX		
CC	Sequence 2, Application US/08460617	
CC	Patent No. 5914114	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Cassels, Frederick J	
CC	TITLE OF INVENTION: Method of Raising Antibodies Against E. E.	
CC	TITLE OF INVENTION: Coll	
CC	NUMBER OF SEQUENCES: 9	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Hendricks and Associates	
CC	STREET: 9669 A Main Street	
CC	CITY: Fairfax	
CC	STATE: VA	
CC	COUNTRY: US	
CC	ZIP: 22031	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Floppy disk	
CC	COMPUTER: IBM PC compatible	
CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: US/08/460.617	
CC	FILING DATE: 02-JUN-1995	
CC	CLASSIFICATION: 42A	

US-08-905-046-2.ra1

Fri Jun 16 16:18:01 2000

```
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hendricks, Glenna
CC REGISTRATION NUMBER: 32,535
CC REFERENCE/DOCKET NUMBER: cas460
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-425-4250
CC TELEFAX: 703-425-2767
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 36 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: E coli
CC SEQUENCE 36 AA; 3629 MW; 7132 CN;

Query Match 91.7%; Score 44; DB 2; Length 36;
Best Local Similarity 75.0%; Pred. No. 5.85e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 27 ASVALTYS 34
   :|||||
QY 1 SAVALTYS 8

RESULT 7
ID PCT-US96-08730-4 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX
DT
DE
DE
XX
XX
Sequence 4, Application PC/TUS9608730
Sequence 4, Application PC/TUS9608730
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick
APPLICANT: Anderson, Jeffrey
APPLICANT: Carter, John Mark
TITLE OF INVENTION: Methods of Raising Antibodies Against E.
TITLE OF INVENTION: Coli of the Family CSF-CFA./1
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glenna Hendricks
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: USA
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08730
FILING DATE: 03-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: PCT/US96/08730
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 591-4470
TELEFAX: (703) 591-4428
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids

CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: internal
CC SEQUENCE 37 AA; 3745 MW; 8426 CN;

Query Match 89.6%; Score 43; DB 4; Length 37;
Best Local Similarity 75.0%; Pred. No. 7.78e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 27 NSVALTYS 34
   :|||||
QY 1 SAVALTYS 8

RESULT 8
ID US-08-460-617-5 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX
DT
DE
DE
XX
XX
Sequence 5, Application US/08460617
Sequence 5, Application US/08460617
Patent No. 5914114
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick J
TITLE OF INVENTION: Method of Raising Antibodies Against E.
TITLE OF INVENTION: Coli
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hendricks and Associates
STREET: 9669 A Main Street
CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,617
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: cas460
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-425-4250
TELEFAX: 703-425-2767
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: E coli
SEQUENCE 37 AA; 3746 MW; 8426 CN;

Query Match 89.6%; Score 43; DB 2; Length 37;
Best Local Similarity 75.0%; Pred. No. 7.78e+01;
```


Fri Jun 16 16:18:01 2000

CC CITY: Fairfax
CC STATE: VA
CC COUNTRY: US
CC ZIP: 22031
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/460,617
CC FILING DATE: 02-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hendricks, Glenna
CC REGISTRATION NUMBER: 32,535
CC REFERENCE/DOCKET NUMBER: cas460
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-425-4250
CC TELEFAX: 703-425-2767
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORGANISM: E coli
CC SEQUENCE 37 AA: 3864 MW; 7776 CN;
Query Match 87.5%; Score 42; DB 2; Length 37;
Best Local Similarity 75.0%; Pred. No. 1.03e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 27 TAVELTYS 34
QY 1 SAVALTYS 8
RESULT 12
ID PCT-US96-08730-5 STANDARD; PRT: 37 AA.
XX
XX
XX
XX
Sequence 5, Application PC/TUS9608730
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick
APPLICANT: Anderson, Jeffrey
APPLICANT: Carter, John Mark
TITLE OF INVENTION: Methods of Raising Antibodies Against E.
TITLE OF INVENTION: Coli of the Family CSF-CFA./1
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glenna Hendricks
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: USA
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08730
FILING DATE: 03-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: PCT/US96/08730
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 591-4470
TELEFAX: (703) 591-4428
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
SEQUENCE 37 AA: 3864 MW; 7776 CN;
Query Match 87.5%; Score 42; DB 4; Length 37;
Best Local Similarity 75.0%; Pred. No. 1.03e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 27 TAVELTYS 34
QY 1 SAVALTYS 8
RESULT 13
ID PCT-US96-08730-11 STANDARD; PRT: 117 AA.
XX
XX
XX
XX
XX
XX
Sequence 11, Application PC/TUS9608730
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick
APPLICANT: Anderson, Jeffrey
APPLICANT: Carter, John Mark
TITLE OF INVENTION: Methods of Raising Antibodies Against E.
TITLE OF INVENTION: Coli of the Family CSF-CFA./1
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glenna Hendricks
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: USA
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08730
FILING DATE: 03-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: PCT/US96/08730
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 591-4470
TELEFAX: (703) 591-4428
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:

CC APPLICATION NUMBER: PCT/US96/08730
CC FILING DATE: 03-JUN-1996
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hendricks, Glenna
CC REGISTRATION NUMBER: 32,535
CC REFERENCE/DOCKET NUMBER: PCT/US96/08730
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 591-4470
CC TELEFAX: (703) 591-4428
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: internal
CC SEQUENCE 37 AA: 3864 MW; 7776 CN;
Query Match 87.5%; Score 42; DB 4; Length 37;
Best Local Similarity 75.0%; Pred. No. 1.03e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 27 TAVELTYS 34
QY 1 SAVALTYS 8
RESULT 13
ID PCT-US96-08730-11 STANDARD; PRT: 117 AA.
XX
XX
XX
XX
XX
XX
Sequence 11, Application PC/TUS9608730
GENERAL INFORMATION:
APPLICANT: Cassels, Frederick
APPLICANT: Anderson, Jeffrey
APPLICANT: Carter, John Mark
TITLE OF INVENTION: Methods of Raising Antibodies Against E.
TITLE OF INVENTION: Coli of the Family CSF-CFA./1
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glenna Hendricks
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: USA
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08730
FILING DATE: 03-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna
REGISTRATION NUMBER: 32,535
REFERENCE/DOCKET NUMBER: PCT/US96/08730
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 591-4470
TELEFAX: (703) 591-4428
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:

CC LENGTH: 117 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: internal
SQ SEQUENCE 117 AA; 12389 MW; 76297 CN;

Query Match 87.5%; Score 42; DB 4; Length 117;
Best Local Similarity 75.0%; Pred. No. 1.03e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 27 TAVELTYS 34
QY :||| ||||
1 SAVALTYS 8

SULT 14
PCT-US96-08730-6 STANDARD; PRT; 36 AA.
XX xxxxxx
AC
XX
XX
DT
XX
DE Sequence 6, Application PC/TUS9608730

CC GENERAL INFORMATION:
CC APPLICANT: Cassels, Frederick
CC APPLICANT: Anderson, Jeffrey
CC APPLICANT: Carter, John Mark
CC TITLE OF INVENTION: Methods of Raising Antibodies Against E.
CC TITLE OF INVENTION: Coli of the Family CSF-CFA./1
CC NUMBER OF SEQUENCES: 15
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Glenna Hendricks
CC STREET: P.O. Box 2509
CC CITY: Fairfax
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22031
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/08730
CC FILING DATE: 03-JUN-1996
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hendricks, Glenna
CC REGISTRATION NUMBER: 32,535
CC REFERENCE/DOCKET NUMBER: PCT/US96/08730
CC TELEPHONE: (703) 591-4470
CC TELEFAX: (703) 591-4428
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 36 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: internal
SQ SEQUENCE 36 AA; 3801 MW; 7306 CN;

Query Match 83.3%; Score 40; DB 4; Length 36;
Best Local Similarity 71.4%; Pred. No. 1.81e+02;

CC LENGTH: 117 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: internal
SQ SEQUENCE 117 AA; 12389 MW; 76297 CN;

Query Match 87.5%; Score 42; DB 4; Length 117;
Best Local Similarity 75.0%; Pred. No. 1.03e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 27 TAVELTYS 34
QY :||| ||||
1 SAVALTYS 8

SULT 14
PCT-US96-08730-6 STANDARD; PRT; 36 AA.
XX xxxxxx
AC
XX
XX
DT
XX
DE Sequence 6, Application PC/TUS9608730

CC GENERAL INFORMATION:
CC APPLICANT: Cassels, Frederick
CC APPLICANT: Anderson, Jeffrey
CC APPLICANT: Carter, John Mark
CC TITLE OF INVENTION: Methods of Raising Antibodies Against E.
CC TITLE OF INVENTION: Coli of the Family CSF-CFA./1
CC NUMBER OF SEQUENCES: 15
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Glenna Hendricks
CC STREET: P.O. Box 2509
CC CITY: Fairfax
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22031
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/08730
CC FILING DATE: 03-JUN-1996
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hendricks, Glenna
CC REGISTRATION NUMBER: 32,535
CC REFERENCE/DOCKET NUMBER: PCT/US96/08730
CC TELEPHONE: (703) 591-4470
CC TELEFAX: (703) 591-4428
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 36 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: internal
SQ SEQUENCE 36 AA; 3801 MW; 7306 CN;

Query Match 83.3%; Score 40; DB 4; Length 36;
Best Local Similarity 71.4%; Pred. No. 1.81e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 28 SIALTYS 34
QY :||| ||||
2 AVALTYS 8

RESULT 15
ID US-08-460-617-7 STANDARD; PRT; 36 AA.
XX xxxxxx
AC
XX
XX
DT
XX
DE Sequence 7, Application US/08460617

CC Sequence 7, Application US/08460617
CC Patent No. 5914114
CC GENERAL INFORMATION:
CC APPLICANT: Cassels, Frederick J
CC TITLE OF INVENTION: Method of Raising Antibodies Against E.
CC TITLE OF INVENTION: Coli
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Hendricks and Associates
CC STREET: 9669 A Main Street
CC CITY: Fairfax
CC STATE: VA
CC COUNTRY: US
CC ZIP: 22031
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08460,617
CC FILING DATE: 02-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hendricks, Glenna
CC REGISTRATION NUMBER: 32,535
CC REFERENCE/DOCKET NUMBER: cas460
CC TELEPHONE: 703-425-4250
CC TELEFAX: 703-425-2767
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 36 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: E coli
SQ SEQUENCE 36 AA; 3817 MW; 7238 CN;

Query Match 83.3%; Score 40; DB 2; Length 36;
Best Local Similarity 71.4%; Pred. No. 1.81e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 28 SIALTYS 34
QY :||| ||||
2 AVALTYS 8

Search completed: Fri Jun 16 15:51:32 2000
Job time : 6 secs.

This Page Blank (uspto)

W A S E R E H (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

on on: Fri Jun 16 15:50:36 2000; MasPar time 4.83 Seconds
Tabular output not generated. 78.084 Million cell updates/sec

Title: >US-08-905-046-2
Description: (1-8) from US08905046.pep
Perfect Score: 48
Sequence: 1 SAVALTYS 8

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.426; Variance 21.680; scale 0.988

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	43	89.6	171	2	fimbrial protein csoA	1.57e+01
2	42	87.5	544	2	hypothetical protein	2.68e+01
3	42	87.5	669	1	DNA ligase (NAD+) (EC 2.68e+01	2.68e+01
4	41	85.4	183	2	hypothetical protein	4.52e+01
5	41	85.4	250	2	hypothetical protein	4.52e+01
6	41	85.4	331	2	iron (iii) abc transp	4.52e+01
7	41	85.4	549	2	dipeptide ABC transpo	4.52e+01
8	41	85.4	739	2	aspartokinase II - Th	4.52e+01
9	40	83.3	81	2	hypothetical protein	7.54e+01
10	40	83.3	202	2	probable lprE protein	7.54e+01
11	40	83.3	229	2	H+-transporting ATP s	7.54e+01
12	40	83.3	343	2	allantoicase (EC 3.5.	7.54e+01
13	40	83.3	370	2	probable H(+)-transpo	7.54e+01
14	40	83.3	375	2	corticotropin-releasi	7.54e+01
15	40	83.3	412	2	probable Na+/H+ anti	7.54e+01
16	40	83.3	415	2	corticotropin-releasi	7.54e+01
17	40	83.3	415	2	corticotropin-releasi	7.54e+01
18	40	83.3	444	2	corticotropin-recept	7.54e+01
19	40	83.3	479	2	hypothetical protein	7.54e+01
20	40	83.3	701	2	heat shock protein 70	7.54e+01
21	40	83.3	889	2	egg sperm receptor -	7.54e+01
22	40	83.3	1827	1	sucrose alpha-glucosi	7.54e+01
23	39	81.3	127	2	cytochrome P450 prote	1.24e+02

24	39	81.3	159	2	E71721	crossover junction en	1.24e+02
25	39	81.3	173	2	T09868	NADH dehydrogenase (u	1.24e+02
26	39	81.3	225	2	T06205	hypothetical protein	1.24e+02
27	39	81.3	235	2	T03755	hypothetical protein	1.24e+02
28	39	81.3	237	2	S73728	probable lipoprotein	1.24e+02
29	39	81.3	239	2	S09277	sensory rhodopsin I -	1.24e+02
30	39	81.3	248	2	H70511	probable prCA protein	1.24e+02
31	39	81.3	255	2	S72865	hypothetical protein	1.24e+02
32	39	81.3	284	2	E72335	probable 4-hydroxyben	1.24e+02
33	39	81.3	330	2	A71945	ketol-acid reductois	1.24e+02
34	39	81.3	330	1	B64561	ketol-acid reductois	1.24e+02
35	39	81.3	360	1	S04030	ferredoxin--NADP+ red	1.24e+02
36	39	81.3	367	1	W2WL11	E2 protein - human pa	1.24e+02
37	39	81.3	455	1	XYECMB	5-methylcytosine-spec	1.24e+02
38	39	81.3	547	1	ERADF4	60.5K fiber protein -	1.24e+02
39	39	81.3	556	2	A44441	B-cell antigen CD19 p	1.24e+02
40	39	81.3	585	2	T00212	major II secretion pat	1.24e+02
41	39	81.3	702	2	S48753	type II surface protein	1.24e+02
42	39	81.3	727	1	S17854	NADH dehydrogenase (u	1.24e+02
43	39	81.3	837	2	C69200	surface proteinase re	1.24e+02
44	39	81.3	912	2	T01769	hypothetical protein	1.24e+02
45	39	81.3	1199	1	S76549	transcription-repair	1.24e+02

ALIGNMENTS

RESULT 1

ENTRY A41467 #type complete
TITLE fimbrial protein csoA - Escherichia coli plasmid pEU405
ALTERNATE_NAMES CSI pilus major subunit
ORGANISM #formal_name Escherichia coli
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
ACCESSIONS A41467; S19003
REFERENCE A41467
#authors Perez-Casal, J.; Swartley, J.S.; Scott, J.R.
#journal Infect. Immun. (1990) 58:3594-3600
#title Gene encoding the major subunit of CSI pill of human enterotoxigenic Escherichia coli.
#cross-references MUID:91034170
#accession A41467
#molecule_type DNA
#residues 1-171 #label PER
#cross-references GB:M37148; GB:M58550; NID:g145573; PIDN:AAA23596.1; PID:g145574

REFERENCE S19002

#authors Jordi, B.J.A.M.; van Vliet, A.H.M.; Willshaw, G.A.; van der Zeijst, B.A.M.; Gastra, W.
#journal FEMS Microbiol. Lett. (1991) 80:265-270
#title Analysis of the first two genes of the CSI fimbrial operon in human enterotoxigenic Escherichia coli of serotype 0139:H28.

#accession S19003

#status preliminary
#molecule_type DNA
#residues 1-171 #label JOR
#cross-references EMBL:X62879; NID:g41169; PIDN:CAA44673.1; PID:g41171

GENETICS

#gene csoA

#genome plasmid

CLASSIFICATION #superfamily CFAI fimbrial protein

SUMMARY #length 171 #molecular-weight 17542 #checksum 5407

Query Match 89.6%; Score 43; DB 2; Length 171;
Best Local Similarity 75.0%; Pred. No. 1.57e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 50 NSVALTYS 57

QY 1 SAVALTYS 8

RESULT 2

Fri Jun 16 16:18:05 2000

ENTRY S76859 #type complete
TITLE hypothetical protein sll1550 - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
ACCESSIONS S76859
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugiyura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S76859
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-544 #label KAN
#cross-references EMBL:D90917; GB:AB001339; NID:g1653836; PID:d1019504; PID:g1653861
#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS
#start_codon GTG
#summary #length 544 #molecular-weight 58631 #checksum 9527
Query Match 87.5%; Score 42; DB 2; Length 544;
Best Local Similarity 85.7%; Pred. No. 2.68e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 364 TVALTYS 370
QY 2 AVALTYS 8
#cross-references MUID:95350830

RESULT 3
ENTRY S74436 #type complete
TITLE DNA ligase (NAD+) (EC 6.5.1.2) - Synecocystis sp. (strain PCC 6803)
#formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 18-Jun-1999
ACCESSIONS S74436
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugiyura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S74436
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-669 #label KAN
#cross-references EMBL:D90899; GB:AB001339; NID:g1651650; PID:BA016588.1; PID:d1017321; PID:g1651660
#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS
#gene lig
#superfamily polydeoxyribonucleotide synthase (NAD+)
#ligase NAD
#length 669 #molecular-weight 74602 #checksum 9660
SUMMARY
Query Match 87.5%; Score 42; DB 1; Length 669;
Best Local Similarity 85.7%; Pred. No. 2.68e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 118 SAVALTY 124
QY 1 SAVALTY 7
#cross-references MUID:95350830

RESULT 4
ENTRY A64009 #type complete
TITLE hypothetical protein HI0507 - Haemophilus influenzae (strain Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
ACCESSIONS A64009
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, J.L.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
#cross-references MUID:95350830
#accession A64009
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-183 #label TIGR
#cross-references GB:U32732; GB:I42023; NID:g1573480; PIDN:AAC22165.1; PID:g1573488; TIGR:HI0507
CLASSIFICATION #superfamily conserved hypothetical protein HI0507
SUMMARY #length 183 #molecular-weight 20814 #checksum 8952

Query Match 85.4%; Score 41; DB 2; Length 183;
Best Local Similarity 62.5%; Pred. No. 4.52e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 164 SAVALTY 171
QY 1 SAVALTY 8
#cross-references MUID:95350830

RESULT 5
ENTRY B70045 #type complete
TITLE hypothetical protein yypB - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS B70045
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conner, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.; Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,

```

M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kashahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Solido,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
#accession B70045 Bacillus subtilis.
#status preliminary; nucleic acid sequence not shown;
translation not shown

#molecule_type DNA
#residues 1-250 #label KUN
#cross-references GB:Z99121; GB:AL009126; NID:g2635827; PID:el186182;
PID:g2636007
#experimental_source strain 168

GENETICS
#gene yvpB
SUMMARY
#length 250 #molecular-weight 27511 #checksum 864
Query Match 85.4%; Score 41; DB 2; Length 250;
Best Local Similarity 62.5%; Pred. No. 4.52e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 105 ATVPLYS 112
:|:|:|:|
1 SAVALYS 8

RESULT 6
ENTRY #type complete
TITLE iron (iii) abc transporter, permease protein (hemu-2) PAB1535
ORGANISM - Pyrococcus abyssi (strain Orsay)
DATE 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
20-Aug-1999

ACCESSIONS G75035
REFERENCE A75001
#authors anonymous, Genoscope
#submission submitted to the EMBL Data Library, July 1999
#description Pyrococcus abyssi genome sequence: Insights into archaeal
chromosome structure and evolution.
#accession G75035
#status preliminary
#molecule_type DNA
#residues 1-331 #label KAW
#cross-references GB:A248287; GB:AL096836; NID:g5458657;
PID:CA850180.1; PID:e1516078; PID:g5458693
#experimental_source strain Orsay
GENETICS
#gene PAB1535

```

```

SUMMARY #length 331 #molecular-weight 35486 #checksum 3725
Query Match 85.4%; Score 41; DB 2; Length 331;
Best Local Similarity 62.5%; Pred. No. 4.52e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 105 AATLAYS 112
:|:|:|:|
1 SAVALYS 8

RESULT 7
ENTRY #type complete
TITLE dipeptide ABC transporter (dipeptide-binding protein) -
Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
20-Aug-1999
ACCESSIONS G69618; S16651
REFERENCE A69580
#authors Kunitz, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kashahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Solido,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
#accession B70045 Bacillus subtilis.
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-549 #label KUN
#cross-references GB:Z99110; GB:AL009126; NID:g2633472;
PID:CA813153.1; PID:el183316; PID:g2633650
#experimental_source strain 168

REFERENCE S16646
#authors Machipoulos, C.; Mueller, J.P.; Slack, F.J.; Murphy, C.G.;
Patankar, S.; Bukusoglu, G.; Sonenshein, A.L.
#journal Mol. Microbiol. (1991) 5:1903-1913
#title A Bacillus subtilis dipeptide transport system expressed
early during sporulation.
#cross-references MUID:92114768

```

```

#accession S16651
#molecule_type DNA
#residues 7-271, 'K', 273-344, 'P', 346-549 ##label MAT
#cross-references EMBL:X56678; NID:g48802
#note the authors translated the codon GTG for residue 10 as
B, AAA for residue 37 as Leu, ACA for residue 111 as
Tyr, ACC for residue 179 as Arg, CTT for residue 180
as Thr, GAA for residue 181 as Leu, GTT for residue
182 as Glu, and ACA for residue 183 as Val

GENETICS
#gene dppE: dcIAE
#superfamily dipeptide transport protein
SUMMARY
#length 549 #molecular-weight 62579 #checksum 8404

Query Match 85.4%; Score 41; DB 2; Length 549;
Best Local Similarity 62.5%; Pred. No. 4.52e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 394 PAITLAYS 401
:|||||
QY 1 SAVALTYS 8

RESULT 8
ENTRY #type complete
TITLE aspartokinase II - Thermotoga maritima (strain MSB8)
ORGANISM #formal_name Thermotoga maritima
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
11-Jun-1999
ACCESSIONS H72364
REFERENCE A72200
#authors Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek,
J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and
Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession H72364
#status preliminary
#molecule_type DNA
#residues 1-739 ##label ARN
#cross-references GB:AE001729; GB:AE000512; NID:g4981049; PID:g4981061;
TIGR:TM0547
##experimental_source strain MSB8

GENETICS
#gene TM0547
SUMMARY
#length 739 #molecular-weight 81434 #checksum 48

Query Match 85.4%; Score 41; DB 2; Length 739;
Best Local Similarity 62.5%; Pred. No. 4.52e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 493 TAILAYS 500
:|||||
QY 1 SAVALTYS 8

RESULT 9
ENTRY #type complete
TITLE hypothetical protein 1 - fission yeast (Schizosaccharomyces
pombe)
ORGANISM #formal_name Schizosaccharomyces pombe
DATE 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
09-Sep-1997
ACCESSIONS S01410
REFERENCE S01410
#authors Kudla, B.; Persuy, M.A.; Gaillardin, C.; Heslot, H.
#journal Nucleic Acids Res. (1988) 16:8603-8617

#title Construction of an expression vector for the fission yeast
Schizosaccharomyces pombe.
#cross-references MUID:88335610
#accession S01410
#molecule_type DNA
#residues 1-81 ##label KUD
#cross-references EMBL:X07027; NID:g4896; PID:g4897
#note the authors translated the codon CAG for residue 2 as
Glu and GGT for residues 32 and 33 as Val
SUMMARY
#length 81 #molecular-weight 8908 #checksum 1917

Query Match 83.3%; Score 40; DB 2; Length 81;
Best Local Similarity 62.5%; Pred. No. 7.54e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 34 AGVPLTYS 41
:|||||
QY 1 SAVALTYS 8

RESULT 10
ENTRY #type complete
TITLE probable lprE protein - Mycobacterium tuberculosis (strain
H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS D70752
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Basham, D.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Davies, R.; Devlin, K.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrall, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession D70752
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-202 ##label COL
#cross-references GB:Z77137; GB:AL123456; NID:g3261593; PID:e254951;
PID:g1480334
##experimental_source strain H37RV

GENETICS
#gene lprE
SUMMARY
#length 202 #molecular-weight 20442 #checksum 9072

Query Match 83.3%; Score 40; DB 2; Length 202;
Best Local Similarity 71.4%; Pred. No. 7.54e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 169 TVALTYA 175
:|||||
QY 2 AVALTYS 8

RESULT 11
ENTRY #type complete
TITLE H+-transporting ATP synthase (EC 3.6.1.34) protein 6 - sea
urchin (Strongylocentrotus purpuratus) mitochondrion (SGC8)
ORGANISM #formal_name mitochondrion Strongylocentrotus purpuratus
DATE 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change
01-Nov-1996
ACCESSIONS S01505
REFERENCE S01505
#authors H+

```

REFERENCE S01499
#authors Jacobs, H.T.; Elliott, D.J.; Math, V.B.; Farquharson, A.
#journal J. Mol. Biol. (1988) 202:185-217
#title Nucleotide sequence and gene organization of sea urchin mitochondrial DNA.
#cross-references MUID:89011951
#accession S01505
#molecule_type DNA
##residues 1-229 #label JAC
##cross-references EMBL:X12631

GENETICS
#genome mitochondrion
#genetic_code SGC8
#start_codon ATA

CLASSIFICATION #superfamily H+-transporting ATP synthase protein 6
KEYWORDS ATP biosynthesis; hydrolase; mitochondrion
SUMMARY #length 229 #molecular-weight 25556 #checksum 7750

Query Match 83.3%; Score 40; DB 2; Length 229;
Best Local Similarity 62.5%; Pred. No. 7.54e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 100 SNISLAYS 107
1 SAVALTYS 8

RESULT 12
ENTRY JH0442 #type complete
TITLE allantoicase (EC 3.5.3.4) - yeast (*Saccharomyces cerevisiae*)
ALTERNATE_NAMES protein YIR029w
ORGANISM #formal_name *Saccharomyces cerevisiae*
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 06-Feb-1998
ACCESSION JH0442; S48491
REFERENCE JH0442
#authors Yoo, H.S.; Cooper, T.G.
#journal Gene (1991) 104:55-62
#title Sequences of two adjacent genes, one (DAL2) encoding allantoicase and another (DCG1) sensitive to nitrogen-catabolite repression in *Saccharomyces cerevisiae*.
#cross-references MUID:92009196
#accession JH0442
#molecule_type DNA
##residues 1-343 #label YOO
##cross-references GB:M64720; NID:gl71367; PID:gl71368
#accession S48478
REFERENCE S48478
#authors Rowley, K.
#submission submitted to the EMBL Data Library, October 1994
#accession S48491
#molecule_type DNA
##residues 1-343 #label ROW
##cross-references GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763374; MIPS:YIR029w

GENETICS
#gene SGD:DAL2; ALC1
##cross-references SGD:S0001468; MIPS:YIR029w
#map_position 9R
KEYWORDS hydrolase
SUMMARY #length 343 #molecular-weight 38714 #checksum 2397

Query Match 83.3%; Score 40; DB 2; Length 343;
Best Local Similarity 75.0%; Pred. No. 7.54e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 204 GAVALKYS 211
1 SAVALTYS 8

RESULT 13
ENTRY A71214 #type complete
TITLE probable H(+)-transporting ATP synthase subunit C -

Pyrococcus horikoshii
#formal_name *Pyrococcus horikoshii*
14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 26-Aug-1999

ACCESSIONS A71214
REFERENCE Kwarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76

#journal Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.
#cross-references MUID:98344137
#accession A71214
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-370 #label KAW
##cross-references GB:AP000007; NID:g3236134; PID:d1032047; PID:g3258421
##experimental_source strain OT3
##note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS
#gene PHI977
CLASSIFICATION #superfamily H+ transporting ATP synthase C subunit
SUMMARY #length 370 #molecular-weight 42602 #checksum 7092

Query Match 83.3%; Score 40; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 7.54e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 330 AVALTY 335
1 SAVALTYS 7

RESULT 14
ENTRY I38879 #type complete
TITLE corticotropin releasing hormone receptor variant - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change 29-May-1998
ACCESSIONS I38879
REFERENCE I38879
#authors Ross, P.C.; Kostas, C.M.; Ramabhadran, T.V.
#journal Biochem. Biophys. Res. Commun. (1994) 205:1836-1842
#title A variant of the human corticotropin-releasing factor (CRF) receptor: cloning, expression and pharmacology.
#cross-references MUID:95110332
#accession I38879
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-375 #label RES
##cross-references EMBL:U16273; NID:g606973; PID:g606974
SUMMARY #length 375 #molecular-weight 43330 #checksum 7120

Query Match 83.3%; Score 40; DB 2; Length 375;
Best Local Similarity 62.5%; Pred. No. 7.54e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 175 TAVLTYS 182
1 SAVALTYS 8

RESULT 15
ENTRY A72473 #type complete
TITLE Probable Na+/H+ antiporter APE2426 - *Aeropyrum pernix* (strain

Fri Jun 16 16:18:05 2000

```

K1)
#formal_name Aeropyrum pernix
#date 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
#accessions A72473
#reference A72473
#authors Kavarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.;
Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.;
Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
#journal DNA Res. (1999) 6:83-101
#title Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
#cross-references MUID:99310339
#accession A72473
#status preliminary
#molecule_type DNA
#residues 1-412 #label KAW
#cross-references DDBJ:AP000064; NID:g5105945; PIDN:BAA81441.1;
#experimental_source strain K1
GENETICS
#gene APE2426
#summary #length 412 #molecular-weight 43883 #checksum 6482
Query Match 83.3%; Score 40; DB 2; Length 412;
Best Local Similarity 71.4%; Pred. No. 7.54e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 174 TSVALTY 180
QY 1 SAVALTY 7
Search completed: Fri Jun 16 15:50:45 2000
Job time : 9 secs.

```

W P S R E F (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 16 15:49:40 2000; Maspar time 3.26 Seconds
Tabular output not generated. 74.780 Million cell updates/sec

Title: >US-08-905-046-2
Description: (1-8) from US08905046.pap
Perfect Score: 48
Sequence: 1 SAVALTY5 8

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 21.794; Variance 19.705; scale 1.106

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	89.6	171	1	FMS1_ECOLI	CS1 FIMBRIAL SUBUNIT A	5.48e+00
2	42	87.5	415	1	CFR8_SHEEP	CORTICOTROPIN RELEASIN	9.82e+00
3	41	85.4	183	1	YQHA_HAEIN	HYPOTHETICAL PROTEIN H	1.74e+01
4	41	85.4	543	1	DPPE_BACSU	DIPEPTIDE-BINDING PROT	1.74e+01
5	40	83.3	202	1	LPRE_MYCTU	PUTATIVE LIPOPROTEIN L	3.03e+01
6	40	83.3	229	1	ATP6_STRPU	ATP SYNTHASE A CHAIN (3.03e+01
7	40	83.3	343	1	ALC_YEAST	ALLANTOICASE (EC 3.5.3	3.03e+01
8	40	83.3	415	1	CFR8_MOUSE	CORTICOTROPIN RELEASIN	3.03e+01
9	40	83.3	415	1	CFR8_RAT	CORTICOTROPIN RELEASIN	3.03e+01
10	40	83.3	415	1	CFR8_XENLA	CORTICOTROPIN RELEASIN	3.03e+01
11	40	83.3	420	1	CFR8_CHICK	CORTICOTROPIN RELEASIN	3.03e+01
12	40	83.3	444	1	CFR8_HUMAN	CORTICOTROPIN RELEASIN	3.03e+01
13	40	83.3	838	1	OS94_MOUSE	OSMOTIC STRESS PROTEIN	3.03e+01
14	40	83.3	839	1	OS94_HUMAN	OSMOTIC STRESS PROTEIN	3.03e+01
15	40	83.3	840	1	H574_HUMAN	HEAT SHOCK 70 KD PROTE	3.03e+01
16	40	83.3	841	1	H574_MOUSE	HEAT SHOCK 70-RELATED	3.03e+01
17	40	83.3	886	1	H597_STRFN	97 KD HEAT SHOCK PROTE	3.03e+01
18	40	83.3	889	1	H597_STRPU	97 KD HEAT SHOCK PROTE	3.03e+01
19	40	83.3	1826	1	SUIS_HUMAN	SUCRASE-ISOMALTAZE, IN	3.03e+01
20	39	81.3	157	1	RUV6_RICPR	CROSSOVER JUNCTION END	5.23e+01
21	39	81.3	173	1	N06M_ONCMY	NADH-UBIQUINONE OXIDOR	5.23e+01
22	39	81.3	173	1	N06M_SALSA	NADH-UBIQUINONE OXIDOR	5.23e+01
23	39	81.3	239	1	BACS_HALHA	SENSORY RHODOPSIN I (S	5.23e+01

24	39	81.3	330	1	ILVC_HELPY	KETOL-ACID REDUCTOISOM	5.23e+01
25	39	81.3	360	1	FENR_PEA	FERREDOXIN--NADP REDUC	5.23e+01
26	39	81.3	363	1	FENR_VICFA	FERREDOXIN--NADP REDUC	5.23e+01
27	39	81.3	367	1	VE2_HPVI1	REGULATORY PROTEIN E2.	5.23e+01
28	39	81.3	453	1	CARA_TRIVE	CARBAMOYL-PHOSPHATE SY	5.23e+01
29	39	81.3	459	1	MCRB_ECOLI	5-METHYLCYTOSINE-SPECI	5.23e+01
30	39	81.3	547	1	FIB1_ADEA0	FIBER PROTEIN 1	5.23e+01
31	39	81.3	556	1	CD19_HUMAN	B-LYMPHOCYTE ANTIGEN C	5.23e+01
32	39	81.3	727	1	NUAM_HUMAN	NADH-UBIQUINONE OXIDOR	5.23e+01
33	39	81.3	919	1	PMP2_HUMAN	PERIODIC TRYPTOPHAN PR	5.23e+01
34	39	81.3	1199	1	MFD_SYNY3	TRANSCRIPTION-REPAIR C	5.23e+01
35	38	79.2	170	1	FMCI_ECOLI	CFA/I FIMBRIAL SUBUNIT	8.91e+01
36	38	79.2	197	1	YCB7_PSEDE	HYPOTHETICAL 21.2 KD P	8.91e+01
37	38	79.2	277	1	YC23_ANTSP	HYPOTHETICAL 30.3 KD P	8.91e+01
38	38	79.2	437	1	HFE2_HAEIN	MINOR FIMBRIAL SUBUNIT	8.91e+01
39	38	79.2	570	1	HEMA_NDVA	HEMAGGLUTININ-NEURAMIN	8.91e+01
40	38	79.2	571	1	HEMA_NDVM	HEMAGGLUTININ-NEURAMIN	8.91e+01
41	38	79.2	590	1	YM63_YEAST	HYPOTHETICAL 67.7 KD P	8.91e+01
42	38	79.2	616	1	HEMA_NDVO	HEMAGGLUTININ-NEURAMIN	8.91e+01
43	38	79.2	776	1	YLA4_CAEEL	HYPOTHETICAL 86.9 KD P	8.91e+01
44	38	79.2	858	1	HI05_CRIGR	HEAT-SHOCK PROTEIN 105	8.91e+01
45	38	79.2	858	1	HI05_HUMAN	HEAT-SHOCK PROTEIN 105	8.91e+01

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	171 AA.
ID	FMS1_ECOLI			
AC	P25730;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	CS1 FIMBRIAL SUBUNIT A PRECURSOR (CS1 PILIN).			
GN	CSOA OR COOA.			
OS	Escherichia coli.			
OG	Plasmid pDEP23, and Plasmid pEU605.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
[1]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ENTEROTOXIGENIC, SEROTYPE 0139:H28; PLASMID=PDEP23;			
RX	MEDLINE; 91357443.			
RA	Jordi B.J.A.M., van Vliet A.H.M., Willshaw G.A.,			
RA	van der Zeijst B.A.M., Gaastria W.;			
RT	"Analysis of the first two genes of the CS1 fimbrial operon in human			
RT	enterotoxigenic Escherichia coli of serotype 0139:H28.;"			
RL	FEMS Microbiol. Lett. 64:265-270(1991).			
[2]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SEROTYPE 06:H16; PLASMID=PEU605;			
RX	MEDLINE; 91034170.			
RA	Perez-Casal J., Swartley J.S., Scott J.R.;			
RT	"Gene encoding the major subunit of CS1 pili of human enterotoxigenic			
RT	Escherichia coli.;"			
RL	Infect. Immun. 58:3594-3600(1990).			
[3]				
RP	SEQUENCE OF 24-42.			
RC	STRAIN=60R75;			
RX	MEDLINE; 90036735.			
RA	Hall R.H., Maneval D.R. Jr., Collins J.H., Theibert J.L.;			
RA	Levine M.M.;			
RT	"Purification and analysis of colonization factor antigen I, coli			
RT	surface antigen I, and coli surface antigen 3 fimbriae from			
RT	enterotoxigenic Escherichia coli.;"			
RL	J. Bacteriol. 171:6372-6374(1989).			
CC	-I- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING			
CC	FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5			
CC	MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO			
CC	COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.			
CC	-I- INDUCTION: CS1 FIMBRIAE ARE ONLY EXPRESSED IN THE PRESENCE OF THE			
CC	POSITIVE REGULATOR RNS.			
CC	-I- SIMILARITY: TO THE CFA/I FIMBRIAL SUBUNIT B (CFAB).			
CC	-----			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

[illegible]

Query Match	89.6%	Score 43;	DB 1;	Length 171;
Best Local Similarity	75.0%	Pred. NO. 5.48e+00;		
Matches	6;	Conservative	2;	Mismatches 0;
Indels	0;	Gaps	0;	

Db	50	NSVALTYS	57
		:	
Qv	1	SAVALTYS	8

RESULT	2	STANDARD;	PRT;	415 AA.
ID	CRFR-SHEEP			
AC	062772;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	CORTICOTROPIN RELEASING FACTOR RECEPTOR 1 PRECURSOR (CRF-R) (CRF1).			
DN	CRHR1.			
OS	Ovis aries (Sheep).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Caprinae; Ovis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=PIUITARY;			
RX	MEDLINE; 99079328.			
RA	Myers D.A., Trinh J.V., Myers T.R.;			
RT	"Structure and function of the ovine type 1 corticotropin releasing			
RT	factor receptor (CRF1) and a carboxyl-terminal variant.;"			
RL	Mol. Cell. Endocrinol. 144:21-35(1998).			
BL	Mol. Neuroendocrinol. A precursor for corticotropin releasing factor			

FUNCTION: THIS IS A RECEPTOR FOR CARCINOGENIC AND
 SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS
 MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYL CYCLASE.
 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
 SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce>
 or send an email to license@sib-sib.ch).

CC	EMBL; AF054582; AAC08027.1; ..
DR	PRINTS; P000249; GPCRSECRETIN.
DR	PROSITE; PS006649; G_PROTEIN_RECEP_F2.1; FALSE_NEG.
DR	PROSITE; PS006650; G_PROTEIN_RECEP_F2.2; 1.
DR	PFAM; PF00002; 7tm2.1.
KW	G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW	Phosphorylation.
FT	SIGNAL 1 24 POTENTIAL.
FT	CHAIN 25 415 CORTICOTROPIN RELEASING FACTOR RECEPTOR
FT	1.
FT	DOMAIN 25 121 EXTRACELLULAR (POTENTIAL).

TRANSMEM	122	142	1 (POTENTIAL).
DOMAIN	143	151	CYTOPLASMIC (POTENTIAL).
TRANSMEM	152	171	2 (POTENTIAL).
DOMAIN	172	189	EXTRACELLULAR (POTENTIAL).
TRANSMEM	190	213	3 (POTENTIAL).
DOMAIN	214	227	CYTOPLASMIC (POTENTIAL).
TRANSMEM	228	249	4 (POTENTIAL).
DOMAIN	250	268	EXTRACELLULAR (POTENTIAL).
TRANSMEM	269	291	5 (POTENTIAL).
DOMAIN	292	314	CYTOPLASMIC (POTENTIAL).
TRANSMEM	315	334	6 (POTENTIAL).
DOMAIN	335	349	EXTRACELLULAR (POTENTIAL).
TRANSMEM	350	369	7 (POTENTIAL).
DOMAIN	370	415	CYTOPLASMIC (POTENTIAL).
CARBOHYD	38	38	POTENTIAL.
CARBOHYD	45	45	POTENTIAL.
CARBOHYD	51	51	POTENTIAL.
CARBOHYD	78	78	POTENTIAL.
CARBOHYD	90	90	POTENTIAL.
CARBOHYD	98	98	POTENTIAL.
SEQUENCE	415 AA:	47558 MW:	FASFG52DI2B4CDC4 CRC64;

Query Match	87.5%	Score 42;	DB 1;	Length 415;
Best Local Similarity	75.0%	Pred. No. 9.82e+00;		
Matches	6:	Conservative	1;	Mismatches 1;
		Indels		

```

Db      215 TAVULTYS 222
Ov      :|||
        1 SAVALTYS 8

```

RESULT	3	STANDARD;	PRT;	183 AA.
ID	YQHA_HAEIN			
AC	P44010:			
CD	01-NOV-1995	(Rel. 32, Created)		
DT	01-NOV-1995	(Rel. 32, Last sequence update)		
DI	01-OCT-1996	(Rel. 34, Last annotation update)		
DE	HYPOTHETICAL PROTEIN HI0507.			
GN	HI0507.			

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.

[1]	SEQUENCE FROM N.A.
RP	STRAIN-RD / KW20;
RC	MEDLINE; 95350630.
RX	
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA	McKenney K., Shitney G., Fitzhugh W., Fields A., Gocayne J.D.,
RA	Scott J.D., Sullivan R., Liu L.-I., Glodak A., Kelley J.M.,
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA	Venter J.C.;

RT "Whole-genome random sequencing and assembly of *Haemophilus*
RT *influenzae* Rd.";
RL Science 269:496-512(1995).

CC	- -	SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL)
CC	- -	SIMILARITY: TO E.COLI YOHA.

[illegible]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC *****
CC *****
DB* EMRI: U32732: AAC22165.1: -

DR TIGR; HI0507; -.
KW Hypothetical protein; Transmembrane.

FT	TRANSMEM	30	50	POTENTIAL.
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10
11	11	11	11	11
12	12	12	12	12
13	13	13	13	13
14	14	14	14	14
15	15	15	15	15
16	16	16	16	16
17	17	17	17	17
18	18	18	18	18
19	19	19	19	19
20	20	20	20	20
21	21	21	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
25	25	25	25	25
26	26	26	26	26
27	27	27	27	27
28	28	28	28	28
29	29	29	29	29
30	30	30	30	30
31	31	31	31	31
32	32	32	32	32
33	33	33	33	33
34	34	34	34	34
35	35	35	35	35
36	36	36	36	36
37	37	37	37	37
38	38	38	38	38
39	39	39	39	39
40	40	40	40	40
41	41	41	41	41
42	42	42	42	42
43	43	43	43	43
44	44	44	44	44
45	45	45	45	45
46	46	46	46	46
47	47	47	47	47
48	48	48	48	48
49	49	49	49	49
50	50	50	50	50
51	51	51	51	51
52	52	52	52	52
53	53	53	53	53
54	54	54	54	54
55	55	55	55	55
56	56	56	56	56
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
61	61	61	61	61
62	62	62	62	62
63	63	63	63	63
64	64	64	64	64
65	65	65	65	65
66	66	66	66	66
67	67	67	67	67
68	68	68	68	68
69	69	69	69	69
70	70	70	70	70
71	71	71	71	71
72	72	72	72	72
73	73	73	73	73
74	74	74	74	74
75	75	75	75	75
76	76	76	76	76
77	77	77	77	77
78	78	78	78	78
79	79	79	79	79
80	80	80	80	80
81	81	81	81	81
82	82	82	82	82
83	83	83	83	83
84	84	84	84	84
85	85	85	85	85
86	86	86	86	86
87	87	87	87	


```

FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
SQ SEQUENCE 183 AA; 20814 MW; B97E2F35E49B267 CRC64;

Query Match 85.4%; Score 41; DB 1; Length 183;
Best Local Similarity 62.5%; Pred. No. 1.74e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 164 SATALAYT 171
   |||||
QY 1 SAVALTYS 8

RESULT 4
ID DPPE_BACSU STANDARD; PRT; 543 AA.
AC P26906;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIPEPTIDE-BINDING PROTEIN DPPE PRECURSOR.
DE DPPE OR DCIAE.
GN Bacillus subtilis.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 92114768.
RA Mathopoulos C., Mueller J.P., Slack F.J., Murphy C.G., Patankar S.,
RA Bukhosoglu G., Sonenshein A.L.;
RT "A Bacillus subtilis dipeptide transport system expressed early
during sporulation.";
RL Mol. Microbiol. 5:1903-1913(1991).
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR DIPEPTIDES; PROBABLY RESPONSIBLE FOR THE BINDING OF DIPEPTIDES
CC WITH HIGH AFFINITY. IS EXPRESSED TO FACILITATE ADAPTATION TO
CC NUTRIENT DEFICIENCY CONDITIONS, WHICH ALSO INDUCE SPOULATION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID ANCHOR
CC (PROBABLE)
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING SPOULATION.
CC -1- INDUCTION: NUTRIENT DEFICIENCY CONDITIONS.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X56678; CAA40006.1; -
DR PIR; S16651; S16651.
DR HSP; P06202; 1RKM.
DR SUBTILIST; BG10846; DPPE.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
DR PFAM; PF00496; SBP_Dac_5; 1.
KW Peptide transport; Transport; Membrane; Signal; Sporulation;
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 543 DIPEPTIDE-BINDING PROTEIN DPPE.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 543 AA; 61818 MW; 9554DDA6F8FDE528 CRC64;

Query Match 85.4%; Score 41; DB 1; Length 543;
Best Local Similarity 62.5%; Pred. No. 1.74e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 388 PAITLTY 395
   |||||
QY 1 SAVALTYS 8

```

```

RESULT 5
ID LPRE_MYCTU STANDARD; PRT; 202 AA.
AC Q11065;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE LIPOPROTEIN LPRE PRECURSOR.
DE LPRE OR RV1252C OR MTCY50.30.
GN Mycobacterium tuberculosis.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (POTENTIAL).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z77137; CAB00900.1; -
DR TUBERCULIST; RV1252C; -
DR KW Hypothetical protein; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 202 PUTATIVE LIPOPROTEIN LPRE.
FT LIPID 29 29 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 202 AA; 20442 MW; E3DCB415A91DAAFE CRC64;

Query Match 83.3%; Score 40; DB 1; Length 202;
Best Local Similarity 71.4%; Pred. No. 3.03e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 169 TVALTYA 175
   |||||
QY 2 AVALTYS 8

RESULT 6
ID ATP6_STRPU STANDARD; PRT; 229 AA.
AC P15995;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
GN ATP6.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OS Mitochondrion.
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89011951.
RA Jacobs H.T., Elliott D.J., Math V.B., Farquharson A.;

```

RT "Nucleotide sequence and gene organization of sea urchin
 RL mitochondrial DNA.";
 CC J. Mol. Biol. 202:185-217(1988).
 CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS. CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X12631; CAA31157.1; ALT_INIT.
 CC PIR; S01505; S01505.
 CC DR PRINTS; PR00123; ATPASBA.
 CC DR PROSITE; PS00449; ATPASE.A; 1.
 CC DR PFAM; PF00119; ATP-synt_A; 1.
 CC Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
 CC KW SEQUENCE 229 AA; 25556 MW; 4BE7F4872DF5EA7A CRC64;
 CC
 CC Query Match 83.3%; Score 40; DB 1; Length 229;
 CC Best Local Similarity 62.5%; Pred. No. 3.03e+01;
 CC Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 100 SNISLTYS 107
 CC | :||| |
 CC QY 1 SAVALTYS 8
 CC
 CC RESULT 7
 CC ID ALC_YEAST STANDARD; PRT; 343 AA.
 CC AC P23335;
 CC DT 01-MAY-1992 (Rel. 22, Created)
 CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
 CC DT 01-FEB-1996 (Rel. 33, Last annotation update)
 CC DE ALLANTOICASE (EC 3.5.3.4).
 CC GN DAL2 OR ALC1 OR YIR029W.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomycetes.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC RX MEDLINE; 92009196.
 CC RA Yoo H.S., Cooper T.G.;
 CC "Sequences of two adjacent genes, one (DAL2) encoding allantoinase
 CC and another (DCG1) sensitive to nitrogen-catabolite repression in
 CC RT Saccharomyces cerevisiae.";
 CC RL Gene 104:55-62(1991).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC RP MEDLINE; 92206079.
 CC RA Lee F.-J.S., Moss J.;
 CC "Cloning of a Saccharomyces cerevisiae gene encoding a protein
 CC RT homologous to allantoinase of Neurospora crassa.";
 CC RL Yeast 7:993-995(1991).
 CC [3]
 CC SEQUENCE FROM N.A.
 CC RP STRAIN-S288C / AB972;
 CC RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 CC RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 CC RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
 CC RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 CC RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 CC RA Walsh S.V., Whitehead S.;
 CC Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: UTILIZATION OF PURINES AS SECONDARY NITROGEN SOURCES,
 CC WHEN PRIMARY SOURCES ARE LIMITING.
 CC -1- CATALYTIC ACTIVITY: ALLANTOATE + H(2)O = (-)-UREIDOGLYCOLATE +
 CC UREA.
 CC -1- PATHWAY: SECOND STEP IN THE DEGRADATION OF ALLANTOIN (PURINE
 CC CATABOLISM).
 CC -1- INDUCTION: REPRESSED BY NITROGEN.
 CC -1- SIMILARITY: STRONG, TO N-CRASSA ALLANTOICASE.
 CC -1- SIMILARITY: TO UREIDOGLYCOLATE HYDROLASE (DAL3).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M64720; AAA34554.1; -;
 CC DR EMBL; X60460; CAA42994.1; -;
 CC DR EMBL; Z38061; CAA86189.1; -;
 CC DR PIR; JH0442; JH0442.
 CC DR SGD; L0000475; DAL2.
 CC KW Hydrolase; Purine metabolism.
 CC FT CONFLICT 93 93 A -> S (IN REF. 2).
 CC FT CONFLICT 134 135 WV -> SL (IN REF. 2).
 CC FT SEQUENCE 343 AA; 38714 MW; 0F9CB0FBA5EB76F1 CRC64;
 CC
 CC Query Match 83.3%; Score 40; DB 1; Length 343;
 CC Best Local Similarity 75.0%; Pred. No. 3.03e+01;
 CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 204 GAVALKYS 211
 CC | :||| |
 CC QY 1 SAVALTYS 8
 CC
 CC RESULT 8
 CC ID CRFR_MOUSE STANDARD; PRT; 415 AA.
 CC AC P35347;
 CC DT 01-JUN-1994 (Rel. 29, Created)
 CC DT 01-JUN-1994 (Rel. 29, Last sequence update)
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
 CC DE CORTICOTROPIN RELEASING FACTOR RECEPTOR 1 PRECURSOR (CRF-R) (CRF1).
 CC GN CRHR1 OR CRHR.
 CC OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC RC TISSUE-PITUITARY;
 CC RX MEDLINE; 94063063.
 CC RA Vita N., Caput D., Ferrara P.;
 CC "Primary structure and functional expression of mouse pituitary and
 CC RT human brain corticotrophin releasing factor receptors.";
 CC RL FEBS Lett. 335:1-5(1993).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
 CC SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS
 CC MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYL CYCLASE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: BRAIN, PITUITARY GLAND, TESTIS. NOT DETECTED
 CC IN PLACENTA, PERIPHERAL BLOOD, LYMPHOCYTES, KIDNEY AND LIVER.
 CC -1- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

```

-----
CC EMBL: X72305; CAA51053.1; -.
DR FR; S39535; S39535.
DR GCRDB: GCR_0848; -.
DR MGD; MGI:88498; CRHR.
DR PRINTS; PR00249; GPCRSECRETIN.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; 1.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
DR PFAM; PF00002; 7tm2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Phosphorylation.
FT SIGNAL 1 24
FT CHAIN 25 415
FT DOMAIN 25 121
FT TRANSMEM 122 142
FT DOMAIN 143 151
FT TRANSMEM 152 171
FT DOMAIN 172 189
FT TRANSMEM 190 213
FT DOMAIN 214 227
FT TRANSMEM 228 249
FT DOMAIN 250 268
FT TRANSMEM 269 291
FT DOMAIN 292 314
FT TRANSMEM 315 334
FT DOMAIN 335 349
FT TRANSMEM 350 369
FT CARBOHYD 38 38
FT CARBOHYD 45 45
FT CARBOHYD 78 78
FT CARBOHYD 90 90
FT CARBOHYD 98 98
SQ SEQUENCE 415 AA; 47769 MW; 81423BDA6D1CA070 CRC64;

Query Match 83.3%; Score 40; DB 1; Length 415;
Best Local Similarity 62.5%; Pred. No. 3.03e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 215 TAIVLYTS 222
QY 1 SAVALYTS 8

RESULT 9 STANDARD; PRT; 415 AA.
ID CRFR_RAT
C P35353;
T 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CORTICOTROPIN RELEASING FACTOR RECEPTOR 1 PRECURSOR (CRF-R) (CRF1).
GN CRHR1 OR CRHR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RX MEDLINE; 94062694.
RA Perrin M.H., Donaldson C.J., Chen R., Lewis K.A., Vale W.W.;
RT "Cloning and functional expression of a rat brain corticotropin
RT releasing factor (CRF) receptor.";
RL Endocrinology 133:3058-3061(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE; 94099969.
RA Chang C.P., Pearce R.V. II, O'Connell S., Rosenfeld M.G.;
RT "Identification of a seven transmembrane helix receptor for
RT corticotropin-releasing factor and sauvagine in mammalian brain.";
RL Neuron 11:1187-1195(1993).
RN [3]

-----
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE; 96278921.
RA Tsai-Morris C., Buczko E., Geng Y., Gamboa-Pinto A., Dufau M.L.;
RT "The genomic structure of the rat corticotropin releasing factor
RT receptor. A member of the class II G protein-coupled receptors.";
RL J. Biol. Chem. 271:14519-14525(1996).
CC 1- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
CC SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE CYCLASE.
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC 1- TISSUE SPECIFICITY: ENTIRE NERVOUS SYSTEM.
CC 1- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
CC 1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L24096; -; NOT_ANNOTATED_CDS.
CC EMBL; L25438; AA16441.1; -.
CC EMBL; U53498; AAC53519.1; -.
CC EMBL; U53486; AAC53519.1; JOINED.
CC EMBL; U53487; AAC53519.1; JOINED.
CC EMBL; U53488; AAC53519.1; JOINED.
CC EMBL; U53489; AAC53519.1; JOINED.
CC EMBL; U53490; AAC53519.1; JOINED.
CC EMBL; U53491; AAC53519.1; JOINED.
CC EMBL; U53492; AAC53519.1; JOINED.
CC EMBL; U53493; AAC53519.1; JOINED.
CC EMBL; U53494; AAC53519.1; JOINED.
CC EMBL; U53495; AAC53519.1; JOINED.
CC EMBL; U53496; AAC53519.1; JOINED.
CC EMBL; U53497; AAC53519.1; JOINED.
CC GCRDB; GCR_0819; -.
CC GCRDB; GCR_0947; -.
CC GCRDB; GCR_1414; -.
CC GCRDB; GCR_1415; -.
CC GCRDB; GCR_1469; -.
CC PRINTS; PR00249; GPCRSECRETIN.
CC PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; 1.
CC PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
CC PFAM; PF00002; 7tm2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Phosphorylation.
FT SIGNAL 1 24
FT CHAIN 25 415
FT DOMAIN 25 121
FT TRANSMEM 122 142
FT DOMAIN 143 151
FT TRANSMEM 152 171
FT DOMAIN 172 189
FT TRANSMEM 190 213
FT DOMAIN 214 227
FT TRANSMEM 228 249
FT DOMAIN 250 268
FT TRANSMEM 269 291
FT DOMAIN 292 314
FT TRANSMEM 315 334
FT DOMAIN 335 349
FT TRANSMEM 350 369
FT CARBOHYD 38 38
FT CARBOHYD 45 45
FT CARBOHYD 78 78
FT CARBOHYD 90 90
FT CARBOHYD 98 98

```

```

SQ SEQUENCE 415 AA; 47842 MW; 48D6704B31D4C013 CRC64;

Query Match 83.3%; Score 40; DB 1; Length 415;
Best Local Similarity 62.5%; Pred. No. 3.03e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Ddb 215 TAIVLTYS 222
      :|: ||||
QY 1 SAVALTYS 8

RESULT 10
ID ID CRFR_XENLA STANDARD; PRT; 415 AA.
O42602;
AC AC
DT DT 15-JUL-1998 (Rel. 36, Created)
DT DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT DT 15-FEB-2000 (Rel. 39, Last annotation update)
DT DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE DE CORTICOTROPIN RELEASING FACTOR RECEPTOR 1 PRECURSOR (CRF-R) (CRF1).
CRF1.
Xenopus laevis (african clawed frog).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
Batrachia; Anura; Mesobatrachia; Pipidea; Xenopodinae;
Xenopus.
[1]
SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RC MEDLINE; 97465573.
RX Dautzenberg F.M., Dietrich K., Palchaudhuri M.R., Spiess J.;
RT "Identification of two corticotropin-releasing factor receptors from
RT Xenopus laevis with high ligand selectivity: unusual pharmacology of
RT the type 1 receptor.";
RT J. Neurochem. 69:1640-1649(1997).
RL -1- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
CC -1- SHOWS HIGH-AFFINITY BINDING FOR UROTENSIN I. THE ACTIVITY OF THIS
CC RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLIL
CC CYCLASE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y14036; CNA74363.1; --
CC GCRDB; GCR_2584; --
CC PRINTS; PR00249; GPCRSECRETIN.
DR PROSITE; PS00649; G_PROTEIN_RECEPT_F2_1; 1.
DR PROSITE; PS00650; G_PROTEIN_RECEPT_F2_2; 1.
DR PFAM; PF00002; 7tm2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 415
CORCOTROPIN RELEASING FACTOR RECEPTOR
1.
DOMAIN 26 121
TRANSMEM 122 142
DOMAIN 143 151
TRANSMEM 152 171
DOMAIN 172 189
TRANSMEM 190 213
DOMAIN 214 227
TRANSMEM 228 250
DOMAIN 250 268
TRANSMEM 269 291
DOMAIN 292 314
TRANSMEM 315 334
DOMAIN 335 349
TRANSMEM 350 369
DOMAIN 370 415
CARBOHYD 38 38
EXTRACELLULAR (POTENTIAL).
1 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
2 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
3 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
4 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
5 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
6 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
7 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
POTENTIAL.

```

FT CARBOHYD 83 83 POTENTIAL.
FT CARBOHYD 95 95 POTENTIAL.
FT CARBOHYD 103 103 POTENTIAL.
SQ SEQUENCE 420 AA; 48600 MW; 8C5C992925F62316 CRC64;
Query Match 83.3%; Score 40; DB 1; Length 420;
Best Local Similarity 62.5%; Pred. No. 3.03e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 220 TAILVAYS 227
QY 1 SAVALVAYS 8
RESULT 12
ID CRFR_HUMAN STANDARD; PRT; 444 AA.
AC P34998; Q13008;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
E CORTICOTROPIN RELEASING FACTOR RECEPTOR 1 PRECURSOR (CRF-R) (CRF1).
GN CRHRI OR CRHR OR CRFR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PITUITARY;
RX MEDLINE; 94022296.
RA Chen R., Lewis K.A., Perrin M.H., Vale W.W.;
RT "Expression cloning of a human corticotropin-releasing-factor
receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8967-8971(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94063063.
RA Vita N., Laurent P., Lefort S., Chalou P., Lelias J.-M., Kaghad M.,
RA le Fur G., Caput D., Ferrara P.;
RT "Primary structure and functional expression of mouse pituitary and
human brain corticotropin releasing factor receptors.";
RL FEBS Lett. 335:1-5(1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Sakai K., Yamada M., Horiba N., Wakui M., Demura H., Suda T.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (CRF-R3).
RX MEDLINE; 95110332.
RA Ross P.C., Kostas C.M., Ramabhadran T.V.;
RT "A variant of the human corticotropin-releasing factor (CRF)
receptor: cloning, expression and pharmacology.";
RL Biochem. Biophys. Res. Commun. 205:1836-1842(1994).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
CC SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYL CYCLASE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: THREE ISOFORMS, CRF-R1, CRF-R2 AND CRF-R3,
CC ARE PRODUCED AS A RESULT OF ALTERNATIVE SPLICING OF THE SAME GENE.
CC CRF-R1 APPEARS TO BE THE DOMINANT FORM. CRF-R3 DOES NOT BIND
CC TO CRF WITH A HIGH AFFINITY.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE CEREBELLUM,
CC PITUITARY, CEREBRAL CORTEX AND OLFACTORY LOBE.
CC -!- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
EMBL; L23333; AAA35719.1;
DR EMBL; L23332; AAA35718.1;
DR EMBL; X72304; CAA51052.1;
DR EMBL; AF039523; AAC69993.1;
DR EMBL; AF039510; AAC69993.1;
DR EMBL; AF039511; AAC69993.1;
DR EMBL; AF039512; AAC69993.1;
DR EMBL; AF039513; AAC69993.1;
DR EMBL; AF039514; AAC69993.1;
DR EMBL; AF039515; AAC69993.1;
DR EMBL; AF039516; AAC69993.1;
DR EMBL; AF039517; AAC69993.1;
DR EMBL; AF039518; AAC69993.1;
DR EMBL; AF039519; AAC69993.1;
DR EMBL; AF039520; AAC69993.1;
DR EMBL; AF039521; AAC69993.1;
DR EMBL; AF039522; AAC69993.1;
DR EMBL; U16273; AAC50073.1;
DR GCRDB; GCR_0679;
DR GCRDB; GCR_0680;
DR GCRDB; GCR_0846;
DR GCRDB; GCR_2030;
DR MIM; 122561;
DR PRINTS; PR00249; GPCRSECRETIN.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; 1.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
DR PFAM; PF00002; 7tm_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Phosphorylation; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 444
FT 1.
FT DOMAIN 25 121
FT TRANSMEM 122 142
FT DOMAIN 143 180
FT TRANSMEM 181 200
FT DOMAIN 201 218
FT TRANSMEM 219 242
FT DOMAIN 243 256
FT TRANSMEM 257 278
FT DOMAIN 279 297
FT TRANSMEM 298 320
FT DOMAIN 321 343
FT TRANSMEM 344 363
FT DOMAIN 364 378
FT TRANSMEM 379 398
FT DOMAIN 399 444
FT CARBOHYD 38 38
FT CARBOHYD 45 45
FT CARBOHYD 78 78
FT CARBOHYD 90 90
FT CARBOHYD 98 98
FT VARSPIC 41 81
FT VARSPIC 146 174
SQ SEQUENCE 444 AA; 50719 MW; 7221AEFB0E7AA8ED CRC64;
Query Match 83.3%; Score 40; DB 1; Length 444;
Best Local Similarity 62.5%; Pred. No. 3.03e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 244 TAILVAYS 251
QY 1 SAVALVAYS 8
RESULT 13
ID OS94_MOUSE STANDARD; PRT; 838 AA.
AC P48722;
DT 01-FEB-1996 (Rel. 33, Created)

Fri Jun 16 16:18:06 2000

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DT OSMOTIC STRESS PROTEIN 94 (HEAT SHOCK 70-RELATED PROTEIN APG-1).
GN OSP94.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96218151.
RA Kojima R., Randall J., Brenner B.M., Gullans S.R.;
RT "Osmotic stress protein 94 (Osp94). A new member of the Hsp110/SSE
RT gene subfamily.";
RL J. Biol. Chem. 271:12327-12332(1996).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN-DDY/STD; TISSUE=TESTIS;
CC MEDLINE: 97160564.
CC Kaneko Y., Nishiyama H., Nonoguchi K., Higashitsuji H., Kishishita M.,
CC Fujita J.;
CC "A novel hsp110-related gene, apg-1, that is abundantly expressed in
CC the testis responds to a low temperature heat shock rather than the
CC traditional elevated temperatures.";
CC J. Biol. Chem. 272:2640-2645(1997).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS.
CC -1- INDUCTION: BY HYPEROSMOLAR SALT STRESS.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U23921; AAC52610.1;
CC EMBL: D49482; BAA08446.1;
CC HSP; P19120; INGC.
CC MGD: MGI:107422; OSP94.
CC PROSITE: PS00297; HSP70_1; FALSE_NEG.
CC PROSITE: PS00329; HSP70_2; 1.
CC PROSITE: PS01036; HSP70_3; 1.
CC PFM: PF00012; HSP70; 1.
CC ATP-binding.
CC CONFLICT 175 176 TA -> HS (IN REF. 2).
CC CONFLICT 221 221 K -> E (IN REF. 2).
CC CONFLICT 279 279 K -> P (IN REF. 2).
CC CONFLICT 308 308 Q -> R (IN REF. 2).
CC CONFLICT 776 776 M -> K (IN REF. 2).
CC SEQUENCE 838 AA; 94385 MW; B2C021DDA7EAF0B1 CRC64;
CC
CC Query Match 83.3%; Score 40; DB 1; Length 838;
CC Best Local Similarity 62.5%; Pred. No. 3.03e+01;
CC Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
CC
CC Db 175 TAVALAYG 182
CC :||||:
CC QY 1 SAVALTYS 8
CC
CC RESULT 14
CC ID OS94 HUMAN STANDARD; PRT; 839 AA.
CC AC O95757;
CC DT 15-FEB-2000 (Rel. 39, Created)
CC DT 15-FEB-2000 (Rel. 39, Last sequence update)
CC DT 15-FEB-2000 (Rel. 39, Last annotation update)
CC DE OSMOTIC STRESS PROTEIN 94 (HEAT SHOCK 70-RELATED PROTEIN APG-1).
CC GN OSP94.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Nonoguchi K., Fujita J.;
RT "Cloning and characterization of human apg-1 and apg-2, members of the
RT hsp110 family, cDNAs and chromosomal assignment of the genes.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB023421; BAA75063.1;
CC PROSITE: PS00297; HSP70_1; FALSE_NEG.
CC PROSITE: PS00329; HSP70_2; 1.
CC PROSITE: PS01036; HSP70_3; 1.
CC ATP-binding.
CC SEQUENCE 839 AA; 94505 MW; AB8D89E8132AAE8E CRC64;
CC
CC Query Match 83.3%; Score 40; DB 1; Length 839;
CC Best Local Similarity 62.5%; Pred. No. 3.03e+01;
CC Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
CC
CC Db 175 TAVALAYG 182
CC :||||:
CC QY 1 SAVALTYS 8
CC
CC RESULT 15
CC ID HS74 HUMAN STANDARD; PRT; 840 AA.
CC AC P34932; O95756;
CC DT 01-FEB-1994 (Rel. 28, Created)
CC DT 15-FEB-2000 (Rel. 39, Last sequence update)
CC DT 15-FEB-2000 (Rel. 39, Last annotation update)
CC DE HEAT SHOCK 70 KD PROTEIN 4 (HEAT SHOCK 70-RELATED PROTEIN APG-2)
CC (HSP70RY).
CC GN HSP4.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC RN [1]
CC SEQUENCE FROM N.A.
CC RA Nonoguchi K., Fujita J.;
CC "Cloning and characterization of human apg-1 and apg-2, members of the
CC hsp110 family, cDNAs and chromosomal assignment of the genes.";
CC Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC [2]
CC SEQUENCE OF 1-749 FROM N.A.
CC RP TISSUE-LYMPHOCYTES;
CC RX MEDLINE: 93329076.
CC RA Fathallah D.M., Cherif D., Dellagi K., Arnaout M.A.;
CC "Molecular cloning of a novel human hsp70 from a B cell line and its
CC assignment to chromosome 5.";
CC J. Immunol. 151:810-813(1993).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB023420; BAA75062.1;
CC EMBL: L12723; AAA02807.1; ALT_FRAME.
CC HSP; P19120; INGC.

```

```

DR MM: 601113;
DR PROSITE; PS00297; HSP70_1; FALSE_NEG.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR PFAM; PF00012; HSP70; 1.
KW ATP-binding; Heat shock; Multigene family.
FT CONFLICT 94 94 L-> W (IN REF. 2).
FT CONFLICT 190 190 A-> R (IN REF. 2).
FT CONFLICT 583 586 NQLL -> ESAI (IN REF. 2).
FT CONFLICT 622 622 R -> E (IN REF. 2).
FT CONFLICT 644 644 G -> D (IN REF. 2).
SQ SEQUENCE 840 AA; 94299 MW; B883AD388E972C4F CRC64;

Query Match 83.38; Score 40; DB 1; Length 840;
Best Local Similarity 62.5%; Pred.No. 3.03e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 175 TAVALAYG 182
:||||:
1 SAVALTYS 8

```

Search completed: Fri Jun 16 15:49:47 2000
Job time : 7 secs.

This Page Blank (uspto)

 W P E R E L I
 ***** (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

On: Fri Jun 16 15:50:05 2000; MasPar time 8.10 Seconds
 Parallel output not generated. 68.471 Million cell updates/sec

Title: >US-08-905-046-2
 Description: (1-8) from US08905046.pep
 Perfect Score: 48
 Sequence: 1 SAVALTYS 8

Scoring table: PAM 150
 Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sprembl12
 1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
 5:sp.invertebrate 6:sp.mammal 7:sp.mnc 8:sp.organelle
 9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
 13:sp.vertebrate 14:sp.virus

Statistics: Mean 20.940; Variance 21.070; scale 0.994

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
1	45	93.8	571 14	073490 HEMAGGLUTININ-NEURAMIN	5.41e+00
2	45	93.8	571 14	093161 HEMAGGLUTININ-NEURAMIN	5.41e+00
3	42	87.5	134 2	047122 CSFA PROTEIN (FRAGMENT	2.86e+01
4	42	87.5	544 2	P74655 HYPOTHETICAL 58.6 KD P	2.86e+01
5	42	87.5	669 2	P72588 DNA LIGASE.	2.86e+01
6	41	85.4	205 10	Q9ZU82 T9J23.22 PROTEIN.	4.89e+01
7	41	85.4	246 2	030519 PROTEASOME ALPHA SUBUN	4.89e+01
8	41	85.4	250 2	034735 YVPB.	4.89e+01
9	41	85.4	549 2	034801 DPPE.	4.89e+01
10	41	85.4	739 2	Q9W217 HOMOSERINE DEHYDROGENA	4.89e+01
11	40	83.3	81 3	Q95185 54/1 GENE.	8.27e+01
12	40	83.3	90 6	077677 CORTICOTROPIN RELEASIN	8.27e+01
13	40	83.3	133 2	Q47121 CSFA PROTEIN (FRAGMENT	8.27e+01
14	40	83.3	135 2	Q47120 CSFA PROTEIN (FRAGMENT	8.27e+01
15	40	83.3	208 2	Q53581 HYPOTHETICAL PROTEIN (8.27e+01
16	40	83.3	370 1	Q57726 370AA LONG HYPOTHETICA	8.27e+01
17	40	83.3	412 1	Q9F959 412AA LONG HYPOTHETICA	8.27e+01
18	40	83.3	454 4	Q99540 LIGASE-LIKE PROTEIN.	8.27e+01
19	40	83.3	474 4	Q14992 HS24/PS2.	8.27e+01
20	40	83.3	479 5	Q17697 SIMILAR TO GLUTAMATE R	8.27e+01

21	40	83.3	509 3	059759 HYPOTHETICAL 57.8 KD P	8.27e+01
22	40	83.3	600 2	P72289 METHYLASE.	8.27e+01
23	40	83.3	662 10	004372 ACID INVERTASE (EC 3.2	8.27e+01
24	40	83.3	693 2	Q56644 HEME RECEPTOR.	8.27e+01
25	40	83.3	840 11	088600 ISCHEMIA RESPONSIVE 94	8.27e+01
26	40	83.3	889 5	Q94769 EGG RECEPTOR FOR SPERM	8.27e+01
27	39	81.3	127 5	Q16913 CYTOCHROME P450 (FRAGM	1.38e+02
28	39	81.3	137 5	Q44540 E03D2.1 PROTEIN.	1.38e+02
29	39	81.3	173 8	Q922M2 NADH DEHYDROGENASE SUB	1.38e+02
30	39	81.3	173 8	Q35251 NADH DEHYDROGENASE SUB	1.38e+02
31	39	81.3	173 8	Q9XN30 NADH DEHYDROGENASE SUB	1.38e+02
32	39	81.3	173 8	Q9XM09 NADH DEHYDROGENASE SUB	1.38e+02
33	39	81.3	235 8	Q35594 ORF-235.	1.38e+02
34	39	81.3	248 2	Q33244 PROTEASOME ALPHA-TYPE	1.38e+02
35	39	81.3	265 2	Q49792 B2126_C3_260.	1.38e+02
36	39	81.3	294 2	Q47375 33KD PROTEIN.	1.38e+02
37	39	81.3	330 2	Q9ZMA9 KETOL-ACID REDUCTOISOM	1.38e+02
38	39	81.3	398 6	Q46639 THYROTROPIN-RELEASING	1.38e+02
39	39	81.3	596 2	Q47423 PLASMID P0157 DNA. PUL	1.38e+02
40	39	81.3	642 2	Q9ZGU0 TYPE II SECRETION PROT	1.38e+02
41	39	81.3	837 1	Q26179 SURFACE PROTEASE RELAT	1.38e+02
42	39	81.3	840 1	Q26190 SURFACE PROTEASE RELAT	1.38e+02
43	39	81.3	1417 14	Q89241 NSP1, NSP2, NSP3 (FRAG	1.38e+02
44	39	81.3	1490 14	Q88791 NONSTRUCTURAL POLYPROT	1.38e+02
45	39	81.3	2590 13	Q9W7R4 TEN-M3.	1.38e+02

ALIGNMENTS

RESULT	1				
ID	073490	PRELIMINARY;	PRT;	571 AA.	
AC	073490;				
DT	01-AUG-1998	(TREMBlrel. 07, Created)			
DT	01-AUG-1998	(TREMBlrel. 07, Last sequence update)			
DT	01-MAY-1999	(TREMBlrel. 10, Last annotation update)			
DE	HEMAGGLUTININ-NEURAMINIDASE.				
GN	HN				
OS	Newcastle disease virus.				
OC	Viruses; ssRNA negative-strand viruses; Mononegavirales;				
OC	Paramyxoviridae; Paramyxovirinae; Rubulavirus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-GAM61;				
RA	GRIBANOV O.G.; STAROV S.K.; SMOLENSKY V.I.; RUDENKO T.V.; DRYGIN V.V.,				
RA	GUSEV A.A.;				
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; Y17261; CAA76719.1; -				
DR	PFAM; PF00423; HN; 1.				
KW	Hemagglutinin.				
SQ	SEQUENCE 571 AA; 62656 MW; AE17E065 CRC32;				
Query Match		93.8%;	Score 45;	DB 14;	Length 571;
Best Local Similarity		87.5%;	Pred. No. 5.41e+00;		
Matches		7;	Conservative 1;	Mismatches 0;	Indels 0; Gaps 0;
Db	40 SAVALTYS 47				
Qy	1 SAVALTYS 8				
RESULT	2				
ID	093161	PRELIMINARY;	PRT;	571 AA.	
AC	093161;				
DT	01-NOV-1998	(TREMBlrel. 08, Created)			
DT	01-NOV-1998	(TREMBlrel. 08, Last sequence update)			
DT	01-MAY-1999	(TREMBlrel. 10, Last annotation update)			
DE	HEMAGGLUTININ-NEURAMINIDASE.				
GN	HN				
OS	Newcastle disease virus (strain H).				
OC	Viruses; ssRNA negative-strand viruses; Mononegavirales;				
OC	Paramyxoviridae; Paramyxovirinae; Rubulavirus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				

RC STRAIN-H;
 RA GRIBANOV O.G., STAROV S.K., SMOLENSKY V.I., RUDENKO T.V., DRYGIN V.V.,
 RA GUSEV A.A.;
 RT "The study of russian strains of Newcastle disease virus.";
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; I17149; CAA76656.1; -
 DR PFAM; PF00423; HN; 1.
 KW Hemagglutinin.
 SQ SEQUENCE 571 AA; 62492 MW; 9743052F CRC32;

Query Match 93.8%; Score 45; DB 14; Length 571;
 Best Local Similarity 87.5%; Pred. No. 5.41e+00;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 40 SAVALAYS 47
 :|||||
 QY 1 SAVALTYS 8

RESULT 3 PRELIMINARY; PRT; 134 AA.
 AC Q47122;
 DT 01-FEB-1997 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
 DE CSFA PROTEIN (FRAGMENT).
 GN CSFA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-9B1373;
 RA GAASTRA W., KUSTERS J.G., VAN DIJK L.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X97493; CAA66124.1; -
 FT NON_TER 1
 FT NON_TER 134
 SQ SEQUENCE 134 AA; 13916 MW; 8BBE8F84 CRC32;

Query Match 87.5%; Score 42; DB 2; Length 134;
 Best Local Similarity 75.0%; Pred. No. 2.86e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 17 TAVELTYS 24
 :|||
 QY 1 SAVALTYS 8

RESULT 4 PRELIMINARY; PRT; 544 AA.
 AC P74655;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE HYPOTHETICAL 58.6 KD PROTEIN.
 OS Synchocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X97493; CAA66124.1; -
 FT NON_TER 1
 FT NON_TER 134
 SQ SEQUENCE 134 AA; 13916 MW; 8BBE8F84 CRC32;

Query Match 87.5%; Score 42; DB 2; Length 134;
 Best Local Similarity 75.0%; Pred. No. 2.86e+01;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90917; BAA18771.1; -
 DR PFAM; PF00395; SLH; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 544 AA; 58631 MW; 027F22B5 CRC32;

Query Match 87.5%; Score 42; DB 2; Length 544;
 Best Local Similarity 85.7%; Pred. No. 2.86e+01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 364 TVALTYS 370
 :|||||
 QY 2 AVALTYS 8

RESULT 5 PRELIMINARY; PRT; 669 AA.
 ID P72588;
 AC P72588;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE DNA LIGASE.
 GN LiG.
 OS Synchocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X97493; CAA66124.1; -
 FT NON_TER 1
 FT NON_TER 134
 SQ SEQUENCE 134 AA; 13916 MW; 8BBE8F84 CRC32;

Query Match 87.5%; Score 42; DB 2; Length 669;
 Best Local Similarity 85.7%; Pred. No. 2.86e+01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 118 SAVALTYS 124
 :|||||
 QY 1 SAVALTYS 7

RESULT 6 PRELIMINARY; PRT; 205 AA.
 ID Q9ZU82;
 AC Q9ZU82;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE T9J23.22 PROTEIN.
 GN T9J23.22
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.

Query Match 87.5%; Score 42; DB 2; Length 669;
 Best Local Similarity 85.7%; Pred. No. 2.86e+01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 118 SAVALTYS 124
 :|||||
 QY 1 SAVALTYS 7

RESULT 6 PRELIMINARY; PRT; 205 AA.
 ID Q9ZU82;
 AC Q9ZU82;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE T9J23.22 PROTEIN.
 GN T9J23.22
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.

Query Match 87.5%; Score 42; DB 2; Length 669;
 Best Local Similarity 85.7%; Pred. No. 2.86e+01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RA LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
RA BARNSTADT M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
RA CARRERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
RA FRASER C.M., VENTER J.C.;
RA "Arabidopsis thaliana chromosome II BAC T9J73 genomic sequence.";
RT Submitted (FEB-1999) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AC006072; AAD13710.1; -
SQ SEQUENCE 205 AA; 22869 MW; F2A9C6F7 CRC32;

Query Match 85.4%; Score 41; DB 10; Length 205;
Best Local Similarity 62.5%; Pred. No. 4.89e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 13 SSIALTYG 20
QY 1 SAVALTYIS 8
I:::||||
1 SAVALTYIS 8

RESULT 7
ID O30519 PRELIMINARY; PRT; 246 AA.
AC O30519;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PROTEASOME ALPHA SUBUNIT.
GN PRCA.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC2-155;
RX KNIPFER N., SHRADER T.E.;
RA MEDLINE; 97426045.
RT "Inactivation of the 20S proteasome in Mycobacterium smegmatis.";
RL Mol. Microbiol. 25:375-383(1997).
DR EMBL; AF009645; AAC45615.1; -
KW Proteasome.
SQ SEQUENCE 246 AA; 26915 MW; 35F40072 CRC32;

Query Match 85.4%; Score 41; DB 2; Length 246;
Best Local Similarity 75.08; Pred. No. 4.89e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 29 SVVALAYS 36
QY 1 SAVALTYIS 8
I:::||||
1 SAVALTYIS 8

RESULT 8
ID O34735 PRELIMINARY; PRT; 250 AA.
AC O34735;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE YVPB.
GN YVPB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC LAZAREVIC V., SOLDI B., RIVOLTA C., REYNOLDS S., MAUEL C.,
RA KARAMATA D.;
RT Submitted (NOV-1997) to the EMBL/GenBank/DBSJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,

```

```

RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUSEPPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
RA KUBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUWANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTELELLA D., PORMOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPAPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEROWSKA A., SEROR S.J., SERRER P., SHIN B.S., SOLDI B.,
RA SORAKI M., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEWAKU K.,
RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZNEGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF017113; AAC67292.1; -
DR EMBL; 299121; CAB15499.1; -
SQ SEQUENCE 250 AA; 27511 MW; 93951575 CRC32;

Query Match 85.4%; Score 41; DB 2; Length 250;
Best Local Similarity 62.5%; Pred. No. 4.89e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 105 ATVPLTYS 112
QY 1 SAVALTYIS 8
I:::||||
1 SAVALTYIS 8

RESULT 9
ID O34801 PRELIMINARY; PRT; 549 AA.
AC O34801;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 12, Last annotation update)
DE DPPE.
GN DPPE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX DEVINE K.M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,

```

RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUNA S., GALIZZI A., GALLERON N.,
RA GHIM S.Y., GLASER P., GORFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPPI G., GUY B.J., HAGA K., HATECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETER P., KONINGSTEIN G., KROGH S., KUNANO M.,
RA KORITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MINUZO M., MOESTL D., NAKAI S., NOBACK M.,
RA NEONE D., O'REILLY M., OGAWA K., OGIWARA K., OUDEGA B., PARK S.H.,
RA PARRO T.M., PORTELELLA D., PORWOLLIG S., PRESCOTT A.M.,
RA PRESCAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADALE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,
RA SOROKIN A., TACCONE E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ002571; CAA05576.1; -;
DR EMBL: Z99110; CAB13153.1; -;
DR HSP: P06202; 1852;
DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
DR PFAM: PF00496; SBP_bac_5; 1.
SQ SEQUENCE 549 AA; 62579 MW; AAC59FF1 CRC32;

Query Match 85.4%; Score 41; DB 2; Length 549;
Best Local Similarity 62.5%; Pred. No. 4.89e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 394 PATILAYS 401
:|:|:|:|
QY 1 SAVALTYS 8

Query Match 83.3%; Score 40; DB 3; Length 81;
Best Local Similarity 62.5%; Pred. No. 8.27e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 34 AGVPLAYS 41
:|:|:|:|
QY 1 SAVALTYS 8

RESULT 10
ID Q9WZ17 PRELIMINARY; PRT; 739 AA.
AC Q9WZ17;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HOMOSERINE DEHYDROGENASE (EC 1.1.1.3) (HDH).
GN TM0547.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99287316.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA McDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RT "Evidence for lateral gene transfer between Archaea and bacteria from
RT genome sequence of *Thermotoga maritima*.";
RL Nature 399:323-329(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,

RA McDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-HOMOSERINE + NAD(P)(+) = L-ASPARTATE BETA-
CC SEMIALDEHYDE + NAD(P)H.
CC -1- PATHWAY: THIRD STEP IN CONVERSION OF L-ASPARTATE TO HOMOSERINE.
CC HOMOSERINE PARTICIPATES IN THE BIOSYNTHESIS OF THREONINE AND THEN
CC ISOLEUCINE AND IN THE BIOSYNTHESIS OF METHIONINE.
CC EMBL: AE001729; AAD35632.1; -;
DR PROSITE: PS01042; HOMOSER_DGHEASE; 1.
KW Kinase; Oxidoreductase; NADP; Threonine biosynthesis;
KW Isoleucine biosynthesis; Methionine biosynthesis.
SQ SEQUENCE 739 AA; 81434 MW; A931A3F3 CRC32;

Query Match 85.4%; Score 41; DB 2; Length 739;
Best Local Similarity 62.5%; Pred. No. 4.89e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 493 TATLAYS 500
:|:|:|:|
QY 1 SAVALTYS 8

RESULT 11
ID Q09185 PRELIMINARY; PRT; 81 AA.
AC Q09185;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE 54/1 GENE.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88335610.
RA KUDLA B., PERSUY M.A., GAILLARDIN C.;
RT "Construction of an expression vector for the fission yeast
RT Schizosaccharomycetes pombe.";
RL Nucleic Acids Res. 16:8603-8617(1988).
DR EMBL: X07027; CAA30076.1; -;
SQ SEQUENCE 81 AA; 8908 MW; 52C180A5 CRC32;

Query Match 83.3%; Score 40; DB 3; Length 81;
Best Local Similarity 62.5%; Pred. No. 8.27e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 34 AGVPLAYS 41
:|:|:|:|
QY 1 SAVALTYS 8

RESULT 12
ID Q77677 PRELIMINARY; PRT; 90 AA.
AC Q77677;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE CORTICOTROPIN RELEASING HORMONE RECEPTOR TYPE I (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PITUITARY;
RA MATTERI R.L., CARROLL J.A.;
RT "Partial cDNA sequence of the porcine corticotropin releasing hormone
RT receptor.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

```

DR  ENBL: AF077185; AAC27320.1; -.
FT  NON_TER      1
FT  NON_TER      90
SQ  SEQUENCE      90 AA; 10594 MW; 295CFCD7 CRC32;

Query Match      83.3%; Score 40; DB 6; Length 90;
Best Local Similarity 62.5%; Pred. No. 8.27e+01;
Matches          5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      51  TAIIVLYS 58
QY      1  SAVALTYS 8

```

[illegible]

```

Query Match      83.38; Score 40; DB 2; Length 133;
Best Local Similarity 71.44; Pred. No. 8.27e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db      18 SIALTYS 24
Qy      2 AVALTYS 8
        :|||||

```

[illegible]

```

Query Match      83.38; Score 40; DB 2; Length 135;
Best Local Similarity 71.45; Pred. No. 8.27e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db      18 SIALTYS 24
Qy      2 AVALTYS 8
       :| | | |

```

```

RESULT 15
ID Q53581 PRELIMINARY; PRT; 208 AA.
AC Q53581;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HYPOTHETICAL PROTEIN (FRAGMENT).
OS Streptomyces albus.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G;
RX MEDLINE; 94299159.
RA CRUZ H., PEREZ C., WELLINGTON E., CASTRO C., SERVIN-GONZALEZ L.;
RT "Sequence of the Streptomyces albus G lipase-encoding gene reveals the
RT presence of a prokaryotic lipase family.";
RL Gene 144:141-142(1994).
DR EMBL; U03114; AAA53486.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 208 AA; 21563 MW; C87D62C9 CRC32;

Query Match 83.3%; Score 40; DB 2; Length 208;
Best Local Similarity 62.5%; Pred. No. 8.27e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 150 AAVALSYG 157
QY 1 SAVALTYS 8
:||||:

```

Search completed: Fri Jun 16 15:50:18 2000
Job time : 13 secs.

This Page Blank (uspto)